DOCUMENT RESUME

ED 376 558 EA 026 244

AUTHOR Luyten, Hans
TITLE Stability of School Effects in Secondary Education: The Impact of Variance across Subjects and Years.
PUB DATE Apr 94
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Analysis of Variance; *Effective Schools Research; Foreign Countries; Longitudinal Studies; Predictor Variables; *School Effectiveness; Secondary Education
IDENTIFIERS Netherlands

ABSTRACT

This paper reports the results of an investigation into the stability of school effects in Dutch secondary education across both years and subjects. The following research questions were addressed: (1) What percentage of the total variance in student achievement per subject can be attributed to differences between schools and to what extent are these effects stable across years? (2) To what extent are school effects stable across subjects? and (3) To what extent does the instability across years interact with the instability across subjects? Methodology involved analysis of datasets provided by the Dutch Ministry of Education and Sciences--the examination results of pre-university track students for the years 1983, 1984, and 1987; senior secondary track students for the years 1983 and 1987; and junior secondary track students for the years 1983, 1984, 1985, and 1987. The school effects per subject were found to be fairly stable across years, but schools appeared to produce remarkably divergent results across subjects. Findings also indicated a substantial interaction effect of instability across years and subjects. The results corroborate the conclusions of recent studies that stressed the important role of departments in secondary schools. The general differences between schools with respect to student achievement were very modest, accounting for no more than 4 percent of the total variance in student achievement. Two figures and one table are included. Contains 43 references. (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 376 558

STABILITY OF SCHOOL EFFECTS IN SECONDARY EDUCATION

The impact of variance across subjects and years

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Luyten

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Hans Luyten

University of Twente, Department of Education

A paper presented to the American Educational Research Association Annual Meeting in New Orleans, Louisiana, April 4-8, 1994

EA 026 244

BEST COPY AVAILABLE

STABILITY OF SCHOOL EFFECTS IN DUTCH SECONDARY EDUCATION

The impact of variance across subjects and years

Hans Luyten, *University of Twente, Department of Education*

This paper reports the results of an investigation into the stability across both years and subjects of school effects in Dutch secondary education. What distinguishes the present study from previous ones dealing with the stability of school effects is the fact that two types of instability have been investigated simultaneously. Not only the instability across years and subjects has been established, but also their interaction. This interaction effect should be interpreted as follows: a school may produce outstanding results with respect to a certain subject one year, while the next year the same school may reveal rather poor results for the same subject. The following specific research questions were addressed:
(1) What percentage of the total variance in student achievement per subject can be attributed to differences between schools and to what extent are these effects stable across years? (2) To what extent are school effects stable across subjects? (3) To what extent does the instability across years interact with the instability across subjects?
The school effects per subject were found to be fairly stable across years, but schools appeared to produce remarkably divergent results across subjects. A substantial interaction effect of instability across years and subjects was detected as well. The findings largely corroborate the conclusions of recent studies stressing the important role of departments in secondary schools. The general differences between schools with respect to student achievement turned out to be very modest, making up no more than 4 % of the total variance in student achievement.

1. STABILITY OF SCHOOL EFFECTS IN THEORY AND RESEARCH

Much research in the field of school effectiveness has been inspired by a strong ambition to direct educational policy (Ralph & Fennesy, 1983). Many authors have been particularly eager to refute the schools-don't-make-a-difference interpretation that was generally attributed to the research outcomes presented by Coleman et al. (1966) and Jencks et al. (1972), even though the general conclusion, stating that the effects of schools on achievement are rather small as compared to the influence of family background, could not be contradicted. As the finding that easily measurable school characteristics like class

size, teacher salaries and experience or the number of books in the library is not consistently related to achievement was corroborated in numerous studies as well, researchers started to focus their attention to the internal functioning of schools. According to Purkey and Smith, however, much of this early school effectiveness literature tended to present "narrow, often simplistic, recipes for school improvement derived from non-experimental data" (Purkey & Smith, 1983; p. 427). Moreover, it was readily assumed that, once the variables causing schools to be more effective were identified, schools could simply decide to change their organizational structure accordingly. At the same time, school level variables that appeared to correlate with high achievement were enthusiastically proclaimed to be causes of school effectiveness. Bossert (1988) has pointed out that a classical, mechanistic model of bureaucratic organization underlies much of the thinking about effective schools. The outcomes of research into the effectiveness of schools have shown that certain features typical of classical bureaucracies coincide with high student achievement. Strong educational leadership, tight coordination and frequent evaluation of pupils' progress emerged as common characteristics of effective schools. In accordance with the conception of schools as classical bureaucracies effectiveness was assumed to be a consistent and stable school characteristic. Hardly any attention was paid to the possibility that a school's effectiveness might vary across grades, classrooms or departments.

The "effective schools model" contrasts sharply with the characterizations of schools as "loosely coupled systems" (Weick, 1976) or as "professional bureaucracies" (Mintzberg, 1979), which suggest that classrooms are isolated workplaces where teachers are quite autonomous in doing their job. Weick has contended that teacher autonomy and loose internal coordination do not entail mere detrimental consequences. Loose coupling might render organizations more flexible, because several autonomous actors within the organization are able to react to changing circumstances in different ways. It should be noted, though, that too much of this flexibility will result in downright chaos. Loosely coupled organizations might also be relatively inexpensive to run, because they require less time and money for coordination. The fact that a loosely coupled system consists of several autonomous units provides considerable room for self-determination by the actors. Mintzberg argues that this professional autonomy impedes rather than stimulates an organization's flexibility. Teachers (and other professionals) generally oppose strict planning and external evaluation of their work, thus making it very difficult for administrators to reform or even control the functioning of professional bureaucracies. Mintzberg's view on the flexibility of schools is more in line with the general experience in the field of educational innovation that schools are hard to change. Both the characterization of schools as loosely coupled systems and professional bureaucracies

depict schools as rather segmented organizations. Descriptions of schools as "a collection of individual entrepreneurs (teachers) surrounded by a common parking lot" or "a group of classrooms held together by a common heating or cooling system" (Murphy, 1992; p. 95) may display the precision of a caricature; accurate in their exaggeration. The possibility that school effectiveness is actually an artefact and that effective schools are simply schools with a high percentage of effective teachers or departments should be taken seriously.

The contrast between the effective schools model on the one hand and the characterization of schools as loosely coupled systems or professional bureaucracies, however, is somewhat artificial, as the effective schools model aims to describe a certain kind of schools, namely the ones with high achieving students, whereas Weick and Mintzberg present a general picture of schools as organizations. The image emerging from school effectiveness research is that the more tightly coordinated schools are the most successful ones. It seems, though, that school improvers inspired by this line of research have not always recognized that ineffective schools might be trapped in a vicious circle: their ineffectiveness may be caused by a lack of internal coordination, which at the same time hampers their ability to change towards a more effective organizational structure. Externally initialized improvement efforts following a top down strategy, which may be suitable in a classical bureaucracy, actually assume the organizational structure to be created already present.

In more recent efforts to construct theoretical models explaining school effectiveness the notion that schools may not be equally effective in all respects at any point in time has been taken into account. Predictors of effectiveness are no longer exclusively school level variables. Explicit attention is paid to variables at several hierarchical levels: classroom and school level, but also higher levels, such as the community, school district and state level (Mortimore et al., 1988; Murphy, 1992; Stringfield & Slavin, 1992). Contingency theory has served as a source of inspiration resulting in the notion that school effectiveness is context-bound. In the models put forward by Purkey and Smith (1983), Scheerens and Creemers (1989) and Scheerens (1990; 1992) classroom instruction is considered to be the basis for school effectiveness. Conditions for effective instruction are constrained or facilitated by organizational conditions, which, in turn, can be constrained or facilitated by environmental conditions. Slater and Teddlie (1992) have addressed the instability of school effectiveness over time in a systematic fashion. Effectiveness is believed to be a function of three major factors: administrative appropriateness, teacher preparedness and student readiness. By treating each factor as a dichotomy schools can be grouped into eight categories or "stages of effectiveness". The most ineffective schools are those scoring low on each factor, whereas the most effective ones score high on each

factor. The six remaining categories can be conceived as intermediate stages between both extremes. Schools are believed to move towards or away from effectiveness along a restricted number of routes.

It follows from these theoretical considerations that effectiveness cannot be assumed to be a stable school characteristic. One and the same school might produce diverging effects in time and within a school both more and less effective teachers and departments will be found. In virtually every study in the field of school effectiveness, however, researchers have had to settle for a rather restricted operationalization of effectiveness. Hardly ever have researchers been able to study a school's effectiveness over a prolonged period of time and comparisons between teachers within schools are relatively scarce as well. To assess student achievement researchers have used either cognitive tests that were quite limited in scope or rather crude attainment measures. If school effects are indeed unstable in certain respects this must have produced some misleading results in a number of instances, because most studies on school effectiveness have dealt with the relationship between student achievement and school characteristics which pertain to the entire school and remain more or less the same in time. Correlations between instable effects and stable school features will mainly reflect coincidental associations between student achievement and general school characteristics. Many of the contradictory findings that have resulted from school effectiveness research might be due to differences in the way effectiveness has been operationalized (Bosker, 1990).

The instability question must be considered one of the major issues in the empirical assessment of school effects together with the adjustment for student background characteristics, test-curriculum overlap and the scope of effectiveness measures. It should be noted, however, that although the question of instability is primarily viewed as a matter of scientific interest, it also has its bearing upon the current debate on market approaches to education, especially the issue of school choice (Levin, 1992). Choosing the right school becomes a very complex decision if school effects are instable in certain respects. Section 2 reviews the findings from previous studies dealing with the (in)stability of school effects. In the remaining paragraphs of the present section the issues of adjustment for student background characteristics, test-curriculum overlap and the scope of effectiveness measures will be briefly discussed.

In school effectiveness research the outcomes of schooling are generally measured by students' scores on cognitive tests. Sometimes so-called "attainment measures" are used, which express the formal educational level pupils have reached after a certain number of years at school (Bosker & Scheerens, 1989). If one wants to establish which school and/or classroom characteristics are related to the academic performance of pupils, it should be taken into account that individual pupil characteristics like general intelligence, previous

achievement and family background usually explain a considerable amount of the variance in academic performance. It is generally acknowledged that in an analysis that seeks to detect school or classroom level variables that can explain a school's effectiveness one should control for such possibly confounding variables. Otherwise differences in academic performance between schools might merely reflect differences in pupil background characteristics.

The use of standardized tests when measuring school effectiveness might still generate a distorted picture even if the scores are adequately controlled for pupil background characteristics, as the content of a test will fit the curriculum of some schools better than others. In many cases it would not be correct to classify a school as ineffective, just because its curriculum does not match a certain test. Only if the test covers topics which every school in the research is required to teach, would such a conclusion be warranted. This might e.g. be the case when the test reflects the educational goals formulated by the government. It should be noted, however, that clearly stated educational goals are more often than not absent, especially when achievement is measured somewhere halfway a long term course. In such cases the test-scores will reflect a school's curricular priorities besides its effectiveness, unless the test-curriculum overlap is adequately controlled for (De Haan, 1992).

The tests employed to measure achievement in school effectiveness research are generally quite limited in scope. Usually either a mathematics, arithmetic or (native) language test is used, so that the outcomes of the analyses only apply to one of these specific aspects of student achievement. Although more than one single measure of student achievement has been taken into account in some studies, analyses in which two or more achievement measures simultaneously serve as criterion variables have hardly ever been performed. The standard procedure in cases where more than one criterion variable is available is to perform a number of separate analyses. The most commonly used techniques of analysis are single-criterion techniques, such as regression analysis, analysis of (co)variance and multi-level analysis (Scheerens, 1992, pp. 51-54/60-64). Sometimes more general indicators are used to measure educational output, e.g. attainment measures expressing the formal level of schooling reached. An analysis of the relationships between independent variables and such a general indicator, however, will yield no more than a fairly crude impression of the relationship between effectiveness and the independent variables. Such an analysis can never reveal whether certain predictor variables are related only to specific aspects of student achievement (e.g. certain subjects) and not to others. The use of a general indicator also obscures the fact that students generally do not perform equally well on every subject.

2. PREVIOUS EMPIRICAL FINDINGS WITH RESPECT TO STABILITY

The amount of research reports mainly or partly dealing with the stability of school effects has accumulated steadily in recent years (Rutter et al., 1979; Rutter, 1983; Gray & Jones, 1985; Cuttance, 1987; Goldstein, 1987; Mandeville & Anderson, 1987; Willms, 1987; Blok & Eiting, 1988; Bosker et al., 1988; Brandsma & Knuver, 1988; Mandeville, 1988; Mortimore et al., 1988; Bosker et al., 1989; Willms & Raudenbush, 1989; Batenburg, 1990; Bosker, 1990; Roeleveld et al., 1990; Bosker, 1991). A summary of the research into the stability of school effects in both primary and secondary education has been presented by Bosker & Scheerens (1989, p. 749). See table 2.1.

TABLE 2.1: *Range of Stability Estimates for School Effects*

	Primary	Secondary
Across years	.35-.65	.70-.95
Across grades	.10-.65	.25-.90
Across classes	.45-1.00*	--
Across subjects	.70-.75	.45-.75
Across criteria	.00-.05	.35-.70

The presented figures are mostly correlation coefficients (pearson's r) expressing the extent to which school effects from two different years, grades, classes, subjects or criteria coincide. Correlations smaller than .70 indicate that more than half of the variance remains to be accounted for. The figures for stability across classes in primary education - marked with an asterisk (*) - represent intra-school correlations (ρ). These figures should be interpreted differently: when ρ is smaller than .50, less than half of the variance is explained. Although Bosker & Scheerens conclude that "school effects do exist even though they may vary across grades, classes, time and criteria" (p. 750) on the basis of these findings, it is also evident that the presented figures also reveal a considerable amount of instability. Values for r consistently larger than .70 were only found between subjects in primary education and between years in secondary education. School effects across classes in primary education seem fairly stable, although ρ-values smaller than .50 have been reported. The figures in table 4.1 corroborate rather than disprove the suspicion

that many of the contradictory findings in school effectiveness research result from instable effect measures.

The extremely low correlations that were reported with respect to the stability across criteria in primary education relate to stability across cognitive and non-cognitive measures of achievement. The stability estimates reported for secondary education refer to correlations between more similar indicators of school effectiveness (e.g. the correlation between the formal level of education reached by students after a number of years and their educational perspectives). Some other interesting findings dealing with the issue of stability across criteria are not reported in table 4.1. The mean achievements per school that are not adjusted for intake differences appear to correlate rather strongly with the unadjusted mean achievements in Dutch primary education, as the (Spearman) rank-correlations range from .78 to .95 (Bosker, 1990; pp. 89-90).

A study into the stability of school effects across subjects in Dutch secondary education has not yet been carried out. The range of stability estimates across subjects in table 2.1 appears to be based on a single study relating to secondary schools in Scotland, in which Cuttance reports a .47 correlation between achievement in English and a general attainment measure, and a .74 correlation between arithmetic achievement and the attainment measure (Cuttance, 1987, pp. 20-21). Willms's findings with respect to stability across subjects in Scottish secondary education are not included in table 2.1. His findings suggest a somewhat stronger stability across subjects for secondary schools in Scotland, as he reports correlations of .69 between general attainment and English, .87 between attainment and arithmetic and .74 between English and arithmetic (Willms, 1987, p. 219). Willms's figures relate to 1980, whereas Cuttance's findings relate to 1981.

Stability of school effects across years in secondary education has been addressed in two Dutch studies (Bosker et al., 1989; Roeleveld et al., 1990), two English studies (Rutter et al., 1979; Goldstein, 1987) and in one Scottish study (Willms, 1987). In all of these studies the school effects were found to be quite stable. However, the interaction of two or more types of instability has never been examined in any systematic fashion. In both Dutch studies, e.g., school effects were assessed using a general attainment measure expressing the formal level of education reached after a certain number of years at school. Two serious drawbacks need to be mentioned regarding the use of these attainment measures. In the first place, the fact that a student's individual achievement varies across subjects, is obscured. Students may get satisfactory results in very different ways. One year the students in a school may get poor results in mathematics and very good ones in English, while it may be the other way round the next year. A researcher using the general attainment measures can never detect such discrepancies between years and would conclude that the school produces a stable output from year to year. Secondly, the

comparability of the attainment measures across schools is questionable, as each school is largely autonomous in deciding whether or not a student is admitted to a higher grade. Only the final examinations are comparable across schools in Dutch secondary education. In section 6 the outcomes of an investigation into the stability of these final examination results are reported. The data originate from 1983 through 1987 and almost the entire range of subjects taught in the secondary schools has been taken into account. The questions the research was intended to answer are listed in the next section.

3. RESEARCH QUESTIONS

The research deals with the stability across years and subjects of the final examination results in Dutch general secondary education. Differences between student background characteristics have been roughly controlled for (see section 5.1). The investigations aimed to answer the following specific questions:

(1) What percentage of the total variance in student achievement can be attributed to differences between schools and to what extent can these school effects be considered to be stable across years? This was investigated separately for (nearly) every examination subject taught in Dutch general secondary education. This question will be dealt with in section 6.1.

(2) To what extent can school effects be considered to be stable across subjects? In other words: are schools only successful in teaching certain subjects or are schools equally successful across the entire range of subjects? See section 6.2.

(3) To what extent do both types of instability, across years and across subjects, interact? It is conceivable that a school appears to be particularly successful with respect to certain subjects in one year, but that the next year the same school presents excellent results with respect to an entirely different set of subjects. See section 6.2.

The outcomes of the analyses are considered to be highly relevant for further theory development and research in the field of school effectiveness, as the explanation of student achievement is the primary goal of both theory and research in this field. The findings will indicate to what extent general characteristics of a school can explain the achievements of its students. If school effects turn out to differ substantially across subjects within schools,

this would suggest that differences between schools are largely attributable to departments within schools. A large amount of instability across years per subject would indicate a strong impact of individual teachers on student achievement. Before describing the datasets and research methods in any detail a general outline of the Dutch system of secondary education is presented in section 4.

4. SECONDARY EDUCATION IN THE NETHERLANDS

The Dutch system of secondary education is subdivided into several curriculum tracks. A major distinction is that between junior vocational training ("LBO") and general secondary education. The junior vocational training consists of several subdivisions. I will not elaborate on this part of the educational system, as the research will focus on the general secondary education. The number of students in the junior vocational training is less than half the number in the general secondary education, which is subdivided into the following three tracks:

- Junior secondary education ("MAVO", 4 year course)
- Senior secondary education ("HAVO", 5 year course)
- Pre-university education ("VWO", 6 year course)

Students are selected for a certain track at the age of twelve on the basis of their (presumed) scholastic aptitude. The advise given by the teacher in the final year of elementary schooling generally plays an important role in the decision which track a certain pupil will follow. There is little mobility between the tracks, but for students who have passed the MAVO examination it is possible to enter into the fourth year of the HAVO course. Students having passed the HAVO examination can enter into the fifth year of the VWO course. Most secondary schools in the Netherlands are single-track schools[1]. These are mainly schools for junior vocational training (LBO) and junior secondary schools (MAVO). Schools that cover the whole range from junior vocational up to pre-university education (from LBO up to VWO) are still relatively scarce. Most multi-track schools only cover a limited range of the entire spectrum of curriculum tracks in Dutch secondary education.

[1]The present government policy, however, is to stimulate the creation of multi-track schools.

The minimum number of final examination subjects for VWO students is seven, for the others it is six. For MAVO students it is possible to take the exam at a higher or a lower level for each subject. Students are allowed to choose which subjects make up their final examination. Dutch and one foreign language, however, are compulsory. Some more detailed limitations apply as well, but these will not be described here. The final examinations for each subject consists of two parts: the school examination ("schoolonderzoek" or SO) and the central written examination ("centraal schriftelijk eindexamen" or CSE). The final grades are established by computing the average of the school examination score and the central written examination score. The school examination consists of at least two tests per subject which are usually developed and administered by the teachers themselves. The assignments for the central written examinations are drawn up by boards established by the minister of education and sciences.

The subjects from which students can choose their examination subjects are not exactly the same for each track. Table 4.1 presents a list of the subjects from which the students in the three separate tracks can choose their examination subjects. These subjects are taught in every school except for Latin and Ancient Greek. Both subjects are taught in the majority of the pre-university schools, however. In some schools other subjects can be chosen as examination subjects as well, e.g. Spanish, Russian, Music or Philosophy, but it is quite exceptional if a subject other than those listed in table 4.1 is chosen as an examination subject. In the pre-university schools the mathematics curriculum is split up into two subjects. Mathematics I deals, (very) roughly speaking, mainly with algebra and Mathematics II mainly with geometry[2].

Teachers work together in departments that coordinate the instruction with respect to one or more subjects. Most departments cover a single subject, the exceptions being the classical languages departments (Latin and Greek), the mathematics departments (Mathematics, Mathematics I and Mathematics II) and the economics departments (General Economics, Business Economics and Economic Awareness). Which subjects a mathematics or an economics department actually deals with depends on the curriculum tracks the school covers. The economics department in a MAVO school, e.g., only deals with economic awareness, whereas the same department in a school covering the VWO, HAVO and MAVO track deals with all three economics subjects. The teachers from departments covering several subjects usually teach every subject their department deals

[2]This situation was changed in 1987, when Mathemathics I and II were substituted for Mathematics A and B. The Mathematics A curriculum has been designed especially for future students in the economic and social sciences, while the Mathematics B curriculum is meant for future students in the natural and technical sciences. The VWO examination results in the present study, however, originate from 1983, 1984 and 1986.

with. The number of teachers belonging to two or more departments is probably very small, although no exact information is available. With respect to the teachers dealing with students in the first grade of general secondary education it has nevertheless been reported that less than 2 % teaches more than a single subject to the same group of students (Matthijssen, 1992; p. 52).

TABLE 4.1: *Examination Subjects in Dutch General Secondary Education*

VWO: pre-university	**HAVO: senior secondary**	**MAVO: junior secondary**
Dutch Language	Dutch Language	Dutch Language
Latin	--	--
Ancient Greek	--	--
French	French	French
German	German	German
English	English	English
History	History	History
Geography	Geography	Geography
--	Mathematics	Mathematics
Mathematics I	--	--
Mathematics II	--	--
Physics	Physics	Physics
Chemistry	Chemistry	Chemistry
Biology	Biology	Biology
General Economics	General Economics	--
Business Economics	Business Economics	--
--	--	Economic Awareness

5. DATA AND METHODS OF ANALYSIS

5.1. Description of the datasets

The analyzed datasets, which were provided by the Dutch ministry of education and sciences, contained information about the examination results in the MAVO track for the years 1983, 1984, 1985 and 1987, about the examination results in the HAVO track for the years 1983 and 1987 and about the examination results in the VWO track for the years 1983, 1984 and 1986. The data with respect to the MAVO and HAVO examination results in 1987 were not complete. The dataset containing the 1987 MAVO examinations included 91% of the students who started the last year of the MAVO course in september 1986 and the HAVO dataset of 1987 contained only 76% of the students who started the last year of their course in 1986. The percentages of "missing students" were much lower in the other years, ranging from about 2.5% to less than 1%. Table 5.1 shows the numbers of schools and students present in the available datasets[3].

TABLE 5.1: *Numbers of schools and students*

Year	VWO:pre-university		HAVO:senior sec.		MAVO:junior sec.		TOTAL	
	Schools	Students	Schools	Students	Schools	Students	Schools	Students
1983	463	35,711	534	52,371	1,101	80,912	1,443	168,994
1984	473	35,421	--	--	1,045	74,404	1,341	109,825
1985	--	--	--	--	1,085	73,305	1,085	73,305
1986	474	36,999	--	--	--	--	474	36,999
1987	--	--	390	38,017	988	65,448	1,143	103,465
Total	499	108,131	570	90,388	1,317	294,069	1,666	492,588

Only the results of the central written examinations were taken into account. The school examination results were not used, as they are not really comparable across schools (Pijl,

[3]One can be sure that some students appear more than once in the datasets. E.g., students that did not pass the examination one year are likely to have tried again the next year. Unfortunately, it was not possible to track down these students. The total numbers of students reported in table 5.1 actually refer to *student records* instead of individual students. The analyses to be reported all deal with these student records. The number of records exceeds the number of individuals with approximately 3 to 4%.

1991). This is also the main reason why the percentage of students per school passing the final examination was not used as a measure for educational output. Whether or not a student passes the final examination is for 50 % determined by the school examination results. Apart from that it is also a rather crude measure which provides no information about the results for the various subjects and only differentiates between students who passed and did not pass. In the case of the MAVO students only the exam scores relating to the higher level examinations were included in the analyses. The average number of available examination scores per MAVO student thus dropped to 5.0, while each student is supposed to take an examination in six subjects. The average number of available scores per HAVO student was 5.9, per VWO student it was 6.9. The HAVO and VWO students are required to take an examination in six and seven subjects respectively.

Because no intake characteristics of the students were available, the research will, strictly speaking, only produce information about the unadjusted achievements of the students. On the other hand, students are selected for a curriculum track on the basis of their scholastic aptitude. The differences between students in one and the same track with respect to their academic capacities may therefore be expected to be relatively limited. It seems justified then to assume that the outcomes of the analyses, which were conducted for each curriculum track separately, present a fairly reliable indication of the school effects and their stability in Dutch secondary education, because differences in intake characteristics have been roughly controlled for. One important consequence of the approach ought yet to be mentioned. It was not possible to take into account information about drop-out or length of the school careers. Schools with high drop-out rates or those that retain their students relatively long before allowing them to go in for the final examination may appear quite effective, although this kind of "effectiveness" clearly contrasts with any common sense image of effective schools. In the present case this is in no sense a serious problem. That would only be the case if the research were aimed at identifying correlates of effectiveness, but the present study focused on the stability of school effects. The reader should realize, however, that when school effects are mentioned, these may also result from high drop-out rates or lengthy school careers.

With respect to the issues of test-curriculum overlap and scope of the effectiveness measures the data leave little to be desired. Almost the entire range of examination subjects is taken into account and the effectiveness measures cover the topics that every school is required to teach, as the examinations reflect the educational goals formulated by the Dutch government. Schools whose curricula do not match the central examinations can be considered as classic examples of ineffectiveness.

5.2. Methods of analysis

The examination results were all standardized per year, subject and curriculum track to z-scores. As a result each examination score was expressed as a deviation from the average score for that particular subject, year and curriculum track. Consider the following (real life) example. One of the VWO students in 1984 got the following scores for Dutch language and Mathematics I: respectively .34 and -.09. This means that in both cases her achievements were quite close to the average scores for those subjects in the 1984 VWO-track. For Dutch language her score was somewhat above and for Mathematics I just below average. Thus comparisons between years and subjects could be more easily made. A disadvantage may be that absolute differences between years or subjects disappear from sight. It is conceivable that students consistently got better results in certain years or for certain subjects. It has been established, however, that the standards employed for computing the central examination scores differ considerably across years (Dutch Education Inspectorate, 1992). Comparisons across years based on unstandardized scores would therefore be meaningless anyway. To compare absolute scores that relate to different subjects and different examinations would at best be a questionable enterprise. It could only show that the *examinations* with respect to certain subjects are more difficult than others, it would not reveal any information about the inherent difficulty of the subjects. The transformation into z-scores still enables the researcher to detect differences between schools and students. Differences between subjects within schools, differences between years within schools and the interactions of year and subject effects within schools can still be detected as well.

The size of school effects per subject and their stability across years was established through multi-level analysis. Using the VARCL-package (Longford, 1986) the total variance in achievement for each subject per curriculum track was partitioned into student level, year level and school level variance. Students were conceived to be nested within years and years within schools.

The question of stability across subjects and its interaction with the stability across years could not be addressed with the help of a multi-level technique, because the available software does not yet provide facilities for dealing with cross-classified levels. Subjects or subject departments can be conceived as nested within schools, but students are not nested within subject departments, as each student takes an examination in several subjects. The year level and the subject level are cross-classified as well. Years cannot be conceived as being nested within subject departments or the other way round. The instability of school effects across subjects and its interaction with instability across years has been assessed by means of an ordinary analysis of variance, in which the schools, the subjects and the years

served as "treatment" variables. The mean school scores per year, per subject were the units of observation. The total variation between these mean school scores could be partitioned into a main school effect, a subject effect, a year effect and an interaction effect of subject by year. Although multi-level analysis provides much better facilities for separating random variance from true parameter variance than an analysis of variance, this drawback is not too serious in the present case as the number of observations per school is fairly large[4].

As each school presented a separate treatment category, the number of treatments largely exceeded the maximum number the available statistical software (SPSS in this case) is designed to handle. A self written computer programme was therefore used to perform the analysis of variance.

6. RESULTS

6.1. School effects per subject and their stability across years

This section deals with the first research question. Several multi-level analyses were carried out in order to compute the size and stability of school effects per track, per subject. For each subject in each track it was established what percentage of the total variance in achievement could be attributed to differences between schools, years and students. Table 6.1 shows only the size of the school effects (expressed as percentages of school level variance) and their stability (expressed as the intra-school between years correlation ρ).

The stability measures in table 6.1 were computed as follows:

$$\rho = \sigma_s^2/(\sigma_s^2+\sigma_y^2)$$

where:

ρ = The stability measure (intra-school between years correlation)
σ_s^2 = The percentage of school level variance
σ_y^2 = The percentage of year level variance

[4]The number of observations per school equals the product of the number of subjects and the number of years. The minimum number of observations per school is 24 and occurs in the HAVO schools (2 years, 12 subjects). The number of observations in the VWO and MAVO schools is 39 and 44 respectively. See also section 6.2.

The table provides sufficient information for computing the percentages of variance at the student or year level. The year level variance can be computed using the formula:

$$\sigma_y^2 = (\sigma_s^2/\rho)-\sigma_\varepsilon^2$$

It is then easy to obtain the percentage of student level variance as the sum of all three percentages (school, year and student level) adds up to one hundred.

Table 6.1 shows that in general no more than 10 % of the variance in student achievement can be attributed to the school level, although this figure varies somewhat across the three curriculum tracks. The size of the school effects differs more seriously across subjects.

On the basis of previous research one would expect the school effects in secondary education to be fairly stable across years (see section 2). The figures in table 6.1 bear out this expectation. Although a substantial amount of year level variance could be observed in most cases, the school level variance generally exceeded this variance across years. Only five out of thirty-eight intra-school correlations turn out to be lower than .50. Four of the five intra-school correlations below .50 were found in the HAVO track. The relatively low stability of effects in the HAVO schools may be due to the fact that these examination results refer to only two years which are rather far apart (1983 and 1987). Apart from these exceptions, however, no serious contradictions were found between the Dutch stability figures per subject and the outcomes based on general attainment measures that have been reported by Bosker et al. (1988) and Roeleveld et al. (1989).

Table 6.1 also shows that the size and stability of the school effects per subject is quite consistent across the three curriculum tracks. Subjects with small school effects in the VWO track display small effects in the other tracks as well and the same can be said about the stability of the effects. Each subject can thus be characterized by the extent to which schools differ with respect to that particular subject and by the extent to which schools produce stable results for that subject. Therefore two scales were constructed. The first one ("Size of school effects") expressing the extent amount of variation between schools for that particular subject, the other one ("Stability of school effects") expressing the stability of this variation. The scale scores were constructed as follows: first the school effects measures and the stability measures were transformed into z-scores per track. Then the average of these z-scores across tracks was computed for each subject. Negative scores thus express that a subject revealed smaller school effects than average or less stable effects than average across the three curriculum tracks. Cronbach's α equals .88 for the "Size of school effects"-scale and .94 for the "Stability of school effects"-scale. Both scales are slightly correlated (r = .23).

TABLE 6.1: *School effects per subject and their stability across years*

Subject	PERCENTAGE OF SCHOOL LEVEL VARIANCE			STABILITY OF SCHOOL LEVEL VARIANCE (ρ)		
	vwo	havo	mavo	vwo	havo	mavo
Dutch Language	5.4 %	5.4 %	5.3 %	.67	.64	.65
Latin	15.5 %	--	--	.71	--	--
Ancient Greek	11.1 %	--	--	.69	--	--
French	9.9 %	6.6 %	13.0 %	.84	.67	.71
German	8.3 %	8.0 %	11.8 %	.88	.88	.87
English	4.9 %	2.4 %	4.2 %	.86	.63	.75
History	8.1 %	4.8 %	10.2 %	.46	.36	.58
Geography	8.3 %	6.4 %	9.6 %	.54	.49	.64
Mathematics	--	10.3 %	15.0 %	--	.62	.68
Mathematics I	6.3 %	--	--	.65	--	--
Mathematics II	7.1 %	--	--	.55	--	--
Physics	6.6 %	6.3 %	8.7 %	.66	.53	.56
Chemistry	7.3 %	7.3 %	11.2 %	.69	.59	.66
Biology	8.1 %	5.4 %	12.2 %	.76	.61	.71
General Economics	6.1 %	4.3 %	--	.53	.36	--
Business Economics	10.4 %	7.3 %	--	.64	.38	--
Economic Awareness	--	--	9.1 %	--	--	.53
Average score across subjects	8.2 %	6.2 %	10.0 %	.67	.56	.67

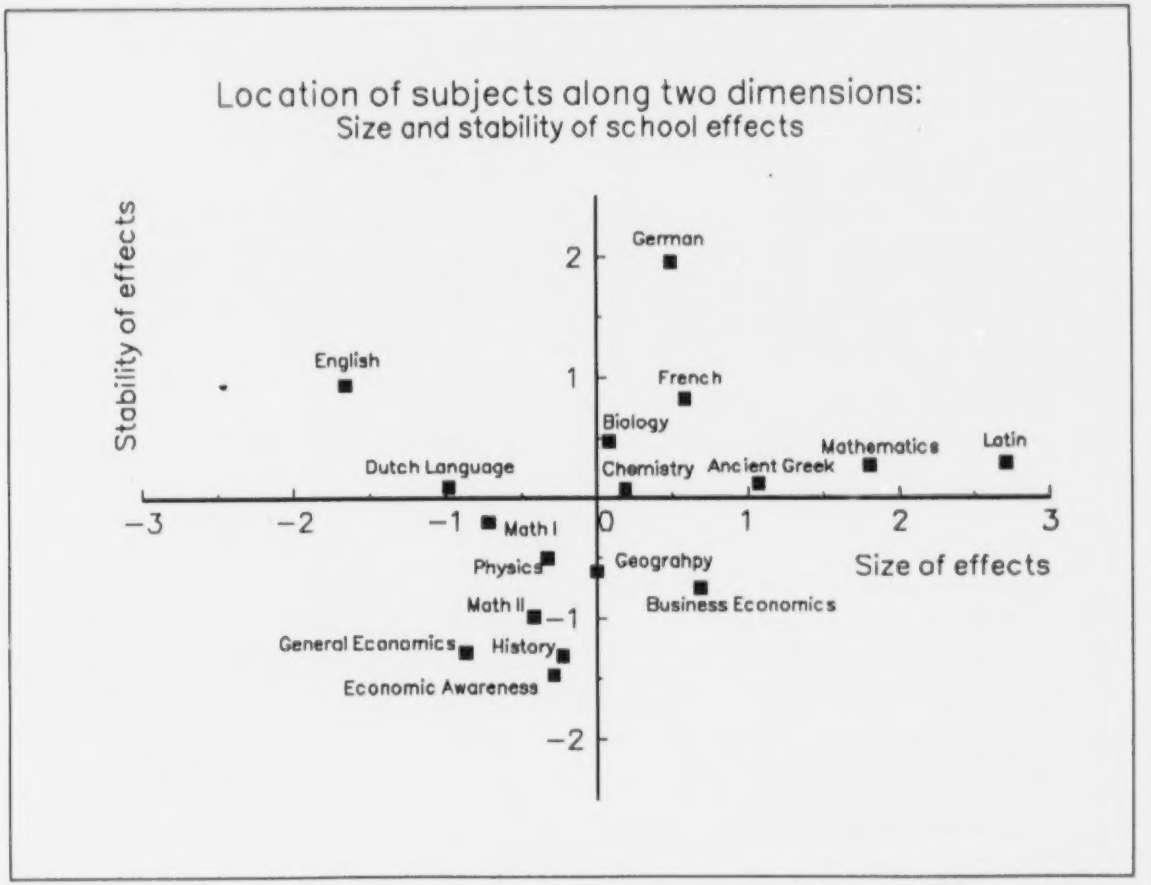
BEST COPY AVAILABLE

In figure 6.1 the subjects are ordered along both dimensions. The figure shows that all language subjects reveal relatively stable effects, as their scores on the stability-scale consistently exceed zero. The economics subjects together with history and geography got generally low scores with respect to stability, while the science and mathematics subjects are somewhere in the middle. The low stability scores for history and geography may be partly due to the fact that the content of the examinations for both these subjects changes from year to year. The internal coordination within departments may explain the stability of effects per subject to some extent as well. Witziers (1992) has found that history departments in Dutch secondary schools are relatively loosely coordinated. His study showed English departments to be much more strongly coordinated. Mathematics departments did also reveal a much stronger internal coordination as compared to the history departments, although not as tight as the internal coordination in the English departments. This coordination mainly concerns the content of instruction, the nature and extent of testing, grading and the goals and outcomes of teaching. Coordination primarily results from joint decision making by the members of the department (Witziers, 1992; pp. 81-98/p. 217).

One would expect to find large school effects for subjects that are predominantly taught within schools. Smaller effects would be expected the more the subjects are learned outside the school as well, e.g. in the case of native language. The research outcomes generally confirm this expectation, but not completely. The findings with respect to the mathematics subjects are particularly remarkable. Whereas mathematics (taught in the HAVO and MAVO track) shows a strong school effect as expected, both mathematics I and II (taught in the VWO track) display relatively small differences between schools. Another surprising result is presented by the small effect for physics.

The remaining subjects do not reveal such unexpected results. The small effect for English language confirms that Dutch children learn much of their English outside the school, especially by watching English language television programmes and by listening to English language pop music. The fact that business economics shows larger effects than the other two economic subjects does not contradict expectations either. Business economics requires more specialized knowledge compared to general economics and economic awareness. Mastering these subjects requires relatively little specialized knowledge, which students need to learn primarily at school and relatively much general knowledge, which may also be acquired elsewhere, e.g. at home. The same can be said about the subjects history and geography.

FIGURE 6.1



21

22

6.2. General school effects and their stability across years and subjects

The outcomes presented in section 6.1 show that the schools produce fairly stable results per subject across years. The present section addresses the two remaining research questions, which relate to the stability across subjects and its interaction with the stability across years. A three-way analysis of variance with one observation per cell was conducted, each cell containing the school mean per year, per subject. The "treatment" variables were the schools, the subjects and the years. Only schools with no missing values for any subject in any year could be included in the analysis, because the applied technique of analysis requires a perfectly balanced design, at least if one wants to partition the total variance into several components (Neter et al., 1985; p. 753). In the case of the VWO track 349 schools were thus included in the analysis. For each of these schools 39 scores (13 subjects, 3 years) were available. Only 13 subject categories were taken into account, because Latin and Greek, which are not taught in a considerable number of schools, were not included in the analysis. Otherwise at least 50 % of the VWO schools should have been excluded[5].

Since the examination scores were transformed into z-scores per year and subject, all mean scores across both years and subjects equalled zero, so that the analysis inevitably revealed zero main effects for subjects, years and for their interaction. No residual variance could be computed because there is only one observation per cell. As a result there are only four kinds of effects to be computed in the analysis: the main school effect, the subject effect within schools, the year effect within schools and the interaction effect of subject by year within schools. The main school effect refers to differences between schools with respect to their mean scores across both years and subjects. E.g., each VWO school has got 39 scores (13 subjects, 3 years). The average of these 39 scores is the mean school score. The subject effect within schools refers to the variation between the subject averages per school. For each VWO school 13 subject averages across years were computed. Each of these subject averages can be expressed as a deviation from the mean school score. The subject effect within schools was computed by summing the squares of these deviations. The year effect within schools refers to the variation between year

[5]Although both Latin and Greek are taught in the majority of the pre-university schools, a large number does not teach these subjects. Latin is taught in less than three quarters of the pre-university schools, while Greek is taught in about 55 % of the VWO schools. Since the applied technique of analysis can only handle cases with no missing values at all (for any subject in any year), including Latin and Greek in the analysis would have resulted in excluding at least 50 % of the pre-university schools.

averages across subjects per school and was computed similarly. The interaction effect of subject by year within schools expresses that a school may be effective with respect to a certain subject one year but much less effective the next year. Together these four effects account for the total variation in all scores.

The analysis of variance was conducted for each curriculum track separately. The results, presented in table 6.2, show that the impact of each effect is roughly the same across the three curriculum tracks. The main school effect appears to constitute only a quarter of the total variance. This means that the impact of general school differences is quite modest in comparison to the joint impact of subject and year effects. The subject effect turns out to be particularly powerful. The interaction effect of subject by year appears to be of equal importance as the main school effect, while the general year effect does not seem very profound. If a general attainment measure had been examined the school effects would have looked much more stable, because in that case only this general year effect and the main school effect could have been detected.

When interpreting the outcomes one should bear in mind that the analysis does not pertain to the individual level, but exclusively to the higher levels of aggregation. From the figures in the bottom row of table 6.1 it can be inferred that at least 85 % of the variance in achievement must be attributed to the individual level. When one combines this information with the outcomes in table 6.2, the message is that school level variables cannot explain more than 4 % of the total variation in achievement in Dutch secondary education, because the main school effect constitutes only 25 % of the remaining 15 %. We should also bear in mind that no information about drop-out or length of the school careers has been taken into account. As a result the main school effect also reflects questionable school policies, such as getting rid of less talented students, or retaining students longer than necessary before allowing them to go in for the final examinations. Besides, differences in intake characteristics have only been roughly controlled for. The main school effect probably reflects such intake differences for some part as well.

The fact that differences between subjects within schools, which are fairly stable themselves, appear to be of more importance than the general school differences should turn our attention to the functioning of departments within secondary schools. It should be borne in mind, however, that the subject effects are certainly not perfectly stable across years and that the interaction effect of subject by year is substantial. Teacher effects seem a very plausible explanation for this interaction. Even though departments generally appear to be fairly tightly coordinated in Dutch secondary education (Witziers, 1992), the impact of individual teachers on student achievement still seems to matter. The magnitude of the teacher effect is probably comparable to that of the main school effect. The general year effect turns out to be very modest. This implies that instability across years affecting

schools with respect to all subjects can explain only a small amount of the variance in student achievement. Instability across years might reflect a widespread but transient organizational disruption within a school or a temporary rise in achievement orientation among the entire teaching staff.

TABLE 6.2: *Three-way analysis of variance on the mean examination results per school School effects, year effects, subject effects and their interactions (expressed as proportional sums of squares)*

	VWO: pre-university (349 schools)	**HAVO: senior secondary** (343 schools)	**MAVO: junior secondary** (639 schools)	**TOTAL**[6] (920 schools)
Main school effect	23.5 %	25.5 %	25.8 %	25.1 %
Main subject effect	set to zero	set to zero	set to zero	set to zero
Main year effect	set to zero	set to zero	set to zero	set to zero
Subject effect within schools	42.6 %	46.5 %	36.5 %	39.8 %
Year effect within schools	6.9 %	7.4 %	8.8 %	8.0 %
Interaction effect of subject by year	set to zero	set to zero	set to zero	set to zero
Interaction effect of subject by year within schools	27.1 %	20.6 %	29.0 %	27.1 %
Residual	cannot be computed	cannot be computed	cannot be computed	cannot be computed

[6]The figures in this column represent the average effects across curriculum tracks. When computing these averages the number of cases per track was taken into account. The number of cases can be computed by multiplying the number of schools with the number of subjects and the number of years. For the VWO track the number of cases is 13,611 (349*13*3). For the HAVO track it is 8,232 (343*12*2) and for the MAVO track 28,116 (639*11*4). Because some schools cover several curriculum tracks, the total number of schools is less than the sum of the VWO schools, HAVO schools and MAVO schools.

FIGURE 6.2

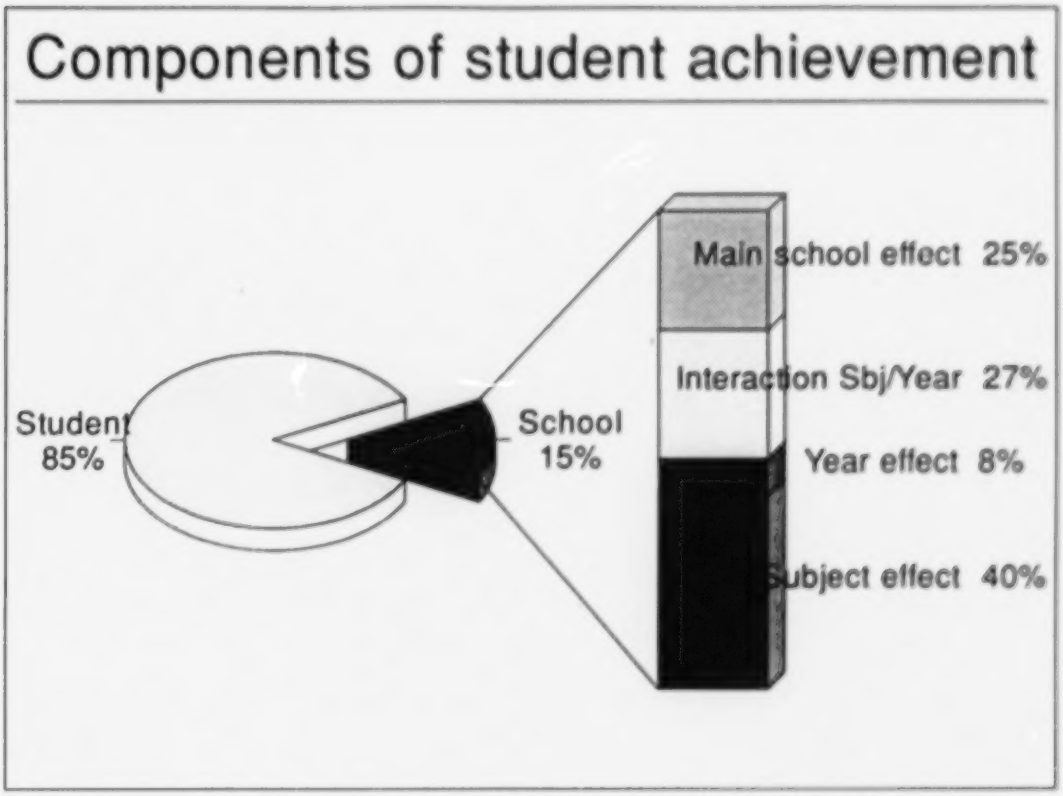


Figure 6.2 presents a graphical description of the relative importance of the main school effect, the subject effect, year effect, and the interaction effect of subject by year on student achievement. It shows that the main school effect represents only the "tip of the iceberg" of what is going on in schools with respect to student achievements. The figure displays that approximately 85 % of the total variance in student achievement is situated at the individual level. No more than a quarter of the remaining variance can be ascribed to main school effects.

7. DISCUSSION

Effectiveness was originally treated as a rather monolithic concept in the thinking about school effectiveness. Hardly any attention was paid to the possibility that within schools certain teachers or departments might be more effective than others or that a school's effectiveness might vary across years. The lack of attention for these kinds of instability can at least partly be explained by the conception of schools as classical bureaucratic organizations that originally underlay much of the thinking about effective schools. Both theoretical considerations and empirical evidence, however, have resulted in the notion that effectiveness is not necessarily a stable school characteristic. The research reported in this paper has shown that school effects in Dutch secondary education do reveal a substantial amount of instability both across subjects and years. The instability across years appeared to interact strongly with the instability across subjects.

The outcomes largely confirm the conclusions of recent studies that departments play an important role in secondary schools (Hylkema, 1990; Siskin, 1991; Witziers, 1992) and that this role should be more thoroughly investigated. Models aiming to explain school effectiveness in secondary education should take into account the impact of departments. The department level should be viewed as an essential intermediate "layer" in the organizational structure of secondary schools, situated between the classroom/teacher level and the school level. In previous research (Hylkema, 1990; Witziers, 1992) the instructional behaviour of Dutch teachers has been shown to be quite strongly regulated through department rules and procedures which result from joint decision making. In this sense teachers appear to restrict their professional autonomy on a voluntary basis. Although the strong interaction effect of subject by year suggests that differences between individual teachers can still have a considerable effect on student achievement, the fact that subject effects make up the largest part of achievement differences between schools implies that the coordination between teachers within departments is stronger than the coordination between departments. Even though departments may be rather tightly coordinated internally, the coordination between them seems to be relatively loose. From this perspective schools for secondary education may still be viewed as loosely coupled systems. However, it is not the individual autonomy of teachers but the collective autonomy of departments that emerges as the main characteristic of secondary schools in the Netherlands.

Future research should pay explicit attention to the internal functioning of departments. It should be investigated to what extent department procedures and regulations affect student achievement and the stability of a department's effectiveness. The available empirical

evidence indicates that tight coordination and rational planning at the department level affects student achievement positively (Witziers, 1992). The present study has shown that history departments, which turned out not to be very tightly coordinated in previous research, are quite instable with respect to their educational output. The assumption that the instability across years per subject mainly reflects teacher effects within departments should be checked, however.

The fact that secondary schools in the Netherlands present such diverging results across years and subjects restricts the opportunities for parents to choose the right school for their children quite seriously. The general differences between schools with respect to achievement turn out to be very modest, as they account for no more 4 % of the total variance in student achievement. Even if parents deliberately choose a school with an outstanding reputation for certain subjects, it still remains to be seen if their children get the right teachers. Although a market approach to education entailing an increased competition among schools for students might stimulate the general effectiveness of secondary education in the long run, it does not seem to make much difference which school parents choose for their children under the present circumstances as far as educational output is concerned.

REFERENCES

Batenburg, Th. A. van (1990). "Variatie in schoolgemiddelden op de Cito-eindtoets basisonderwijs" (Variation in school means on the final elementary education test). *Tijdschrift voor Onderwijsresearch*, **15**, 362-369.

Blok, H. & Eiting, M.H. (1988). "De grootte van schooleffecten: Hoe verschillend presteren leerlingen van verschillende scholen?" (The size of school effects: How different are the achievements of students from different schools?). *Tijdschrift voor Onderwijsresearch*, **13**, 16-30.

Bosker, R.J. (1990). "Theory development in school effectiveness research: in search for consistency of effects". In: P. van den Eeden, J. Hox & J. Hauser (Eds.): *Theory and model in multilevel research: convergence or divergence?* (pp. 77-98) Amsterdam: SISWO.

Bosker, R.J. (1991). "De consistentie van schooleffecten in het basisonderwijs" (The consistency of school effects in elementary education). *Tijdschrift voor Onderwijsresearch*, **16**, 206-218.

Bosker, R.J., Guldemond, H., Hofman, R.H. & Hofman, W.H.A. (1988). *Kwaliteit in het voortgezet onderwijs* (Quality in secondary education). Groningen: RION.

Bosker, R.J., Guldemond, H., Hofman, R.H. & Hofman, W.H.A. (1989). "De stabiliteit van schoolkwaliteit" (The stability of school quality). In: J. Scheerens & J.C. Verhoeven (Eds.): *Schoolorganisatie, beleid en onderwijskwaliteit* (pp. 39-53). Lisse: Swets & Zeitlinger.

Bosker, R.J. & Scheerens, J. (1989). "Issues in the interpretation of the results of school effectiveness research". *International Journal of Educational Research*, **13**, 741-751.

Bossert, S.T. (1988). "School Effects" In: N. Boyan (Ed.), *Handbook of Research on Educational Administration* (pp. 341-352). New York: Longman.

Brandsma, H.P. & Knuver, J.W.M. (1988). "Organisatorische verschillen tussen basisscholen en hun effect op leerlingprestaties" (Organizational differences between elementary schools and their effects on student achievement). *Tijdschrift voor Onderwijsresearch*, **13**, 201-212.

Coleman, J.S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. & York, R. (1966). *Equality of educational opportunity*. Washington, D.C.: Government Printing Office.

Cuttance, P. (1987). *Modelling variation in the effectiveness of schooling*. Edinburgh: CES.

De Haan, D.M. (1992). *Measuring test-curriculum overlap* (Dissertation). Enschede: University of Twente.

Dutch Education Inspectorate (1992). *Examens op punten getoetst* (Examinations tested). The Hague: Ministry of Education and Sciences.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Charles Griffin.

Gray, J. & Jones, B. (1985). "Combining quantitative and qualitative approaches to studies of school and teacher effectiveness. In: D. Reynolds (Ed.): *Studying school effectiveness* (pp. 103-115). London: The Falmer Press.

Hylkema, W.F.S. (1990) *Docenten en hun vaksecties* (Teachers and their subject departments; Dissertation, with an English summary). Nijmegen: Catholic University of Nijmegen

Jencks, C.S., Smith, M., Ackland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B. & Michelson, S. (1972). *Inequality: a reassesment of the effect of family and schooling in America*. New York: Basic.

Levin, H.M. (1992). "Market approaches to education: Vouchers and School Choice" *Economics of Education Review*, **11**, 279-285.

Longford, N.T. (1986). *Variance component analysis: manual*. University of Lancaster

Mandeville, G.K. (1988). "School effectiveness indices revisited: cross-year stability". *Journal of Educational Measurement*, **25**, 349-356.

Mandeville, G.K. & Anderson, L.W. (1987). "The stability of school effectiveness indices across grade levels and subject areas". *Journal of Educational Measurement*, **24**, 203-216.

Matthijssen, Ch. (1992). *Totaalonderwijs, meervoudig bevoegde leraren een uitstervend ras, ervaringen en effecten van het breed inzetten van leraren* (Teachers with multiple qualifications a dwindling race). Enschede: University of Twente, OCTO.

Mintzberg, H. (1979). *The Structuring of Organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988). *The junior school project*. Somerset: Open Books

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988). *The junior school project: technical appendices*. London: ILEA, Research and Statistics Branch.

Murphy, J. (1992). "School effectiveness and school restructuring: contributions to educational improvement" *School Effectiveness and School Improvement*, **3**, 90-109.

Neter, J., Wasserman, W. & Kutner, M.H. (1985) *Applied linear statistical models*(second edition). Homewood, Illinois: Irwin

Pijl, Y.J. (1991) *Discrepanties tussen het schoolonderzoek en het centraal schriftelijk examen in HAVO en VWO* (Discrepancies between the school examination scores and the central written examinations in senior secondary and pre-university education). Groningen: RION.

Purkey, S.C. & Smith, M.S. (1983). "Effective schools: a review". *Elementary School Journal*, **83**, 427-452.

Ralph, J.H. & Fennessy, J. (1983). "Science or reform: some questions about the effective schools model". *Phi Delta Kappan*, **64**, 689-694.

Roeleveld, J., Jong, U., de & Koopman, P. (1990). "De stabiliteit van schooleffecten" (The stability of school effects). *Tijdschrift voor Onderwijsresearch*, **15**, 301-316.

Rutter, M. (1983). "School effects on pupil progress. Research findings and policy implications". *Child Development*, **54**, 1-29.

Rutter, M., Maugham, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours. Secondary schools and their effects on children*. Somerset: Open Books.

Scheerens, J. (1990). "School effectiveness research and the development of process indicators of school functioning" *School Effectiveness and School Improvement*, *1*, 61-80.

Scheerens, J. (1992). *Effective Schooling, research, theory and practice*. London/New York:Cassell.

Scheerens, J. & Creemers, B. (1989). "Towards a more comprehensive conceptualization of school effectiveness". In: B. Creemers, T. Peters & D. Reynolds (eds.): *School Effectiveness and School Improvement* (pp. 265-287). Lisse: Swets & Zeitlinger.

Siskin, L.S. (1991). "Departments as different worlds: Subject cultures in secondary schools". *Educational Administration Quarterly*, *17*, 134-160.

Slater, R.O & Teddlie, C. (1992) "Toward a theory of school effectiveness and leadership". *School Effectiveness and School Improvement*, *3*, 242-257.

Stringfield, S.C. & Slavin, R.E. (1992) "A hierarchical longitudinal model for elemenentary school effect". In: B.P.M. Creemers & G.J. Reezigt (eds.): *Evaluation of Educational Effectiveness* (pp. 35-68). Groningen: ICO

Weick, K. (1976). "Educational organizations as loosely coupled systems". *Administrative Science Quarterly*, **21**, 1-19.

Willms, J.D. (1987). "Differences between Scottish education authorities in their examination attainment". *Oxford Review of Education*, **13**, 211-232.

Willms, J.D. & Raudenbush, S.W. (1989). "A longitudinal hierarchical linear model for estimating school effects and their stability". *Journal of Educational Measurement*, **26**, 209-232.

Witziers, B. (1992). *Coördinatie binnen scholen voor voortgezet onderwijs* (Coordination in secondary schools; Dissertation, with an English summary). Enschede: University of Twente.

BEST COPY AVAILABLE